



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



Publication number:

0 688 870 A1

(2)

## EUROPEAN PATENT APPLICATION

(1) Application number: 95104959.2

(5) Int. Cl<sup>s</sup>: C12N 15/48, C12N 9/12,  
C07K 14/15

(2) Date of filing: 03.04.95

The applicant has subsequently filed a sequence listing and declared, that it includes no new matter.

(3) Priority: 01.04.94 US 221804

(4) Date of publication of application:  
27.12.95 Bulletin 95/52

(6) Designated Contracting States:  
AT BE CH DE DK ES FR GB IT LU NL SE

(7) Applicant: GEN-PROBE INCORPORATED  
9880 Campus Point Drive  
San Diego  
California 92121-1514 (US)

(8) Inventor: Kacian, Daniel Louis

3911 Tambor Road  
San Diego,  
California 92024 (US)  
Inventor: Riggs, Michael Garth  
5715 Lodi Street  
San Diego,  
California 92117 (US)  
Inventor: Putnam, James Garfield  
2104 Dale Street  
San Diego,  
California 92104 (US)

(9) Representative: Patentanwälte Vierling &  
Jentschura  
Postfach 22 14 43  
D-80504 München (DE)

### (10) Highly-purified recombinant reverse transcriptase

(11) A plasmid for expression of Moloney Murine Leukemia Virus-derived reverse transcriptase in *E. coli* cells deficient in the expression of indigenous RNase activity, a method for purification of the recombinant enzyme, and a composition comprising a cloned and purified reverse transcriptase optimized for use in cDNA and nucleic acid amplification procedures.

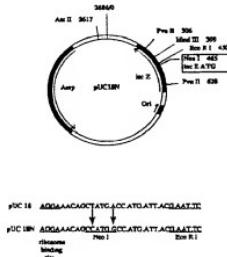


FIG. 1.

EP 0 688 870 A1

Background of the invention

Retroviruses are a group of viruses whose genetic material consists of single-stranded RNA. Following adsorption and entry of the retroviral RNA into the host cell, the viral RNA serves as a template for the synthesis of a complementary DNA strand. The DNA is then made double-stranded through the action of an enzyme having DNA polymerase activity; it is this double-stranded DNA which integrates into the host genome. The RNA-directed DNA polymerase activity responsible for the synthesis of complementary DNA from the viral RNA template is commonly called reverse transcriptase.

- Retroviruses are of particular interest because a number of retroviruses have been implicated as the causative agents of various cancers, and other diseases. A retrovirus, human immunodeficiency virus, is the causal agent of acquired immunodeficiency syndrome (AIDS). Additionally, the reverse transcriptase enzymes themselves have become important reagents in molecular biology because of their ability to make complementary DNA from almost any RNA template. Thus, reverse transcriptase is commonly used to make nucleic acids for hybridization probes and to convert single-stranded RNA into a double-stranded DNA for subsequent cloning and expression.

Recently, reverse transcriptases have been used as a component of transcription-based amplification systems. These systems amplify RNA and DNA target sequences up to 1 trillion fold. See e.g., Burg et al., PCT Patent Application WO 89/01050 (1988); Gingeras et al., PCT Patent Application WO 88/10315 (1988); Davey and Malek, European Patent Application EPO 0329822 (1988); Gingeras et al., European Patent Application EPO 0373960 (1989); Malek and Davey, PCT Patent Application WO 91/02814 (1989); Kacian and Fultz, European Patent Application EPO 0408295 A2 (1990). All of these references are hereby incorporated by reference into this disclosure.

Some of the transcription-based amplification methods are exceptionally convenient since the amplification reaction according to these methods is isothermal. Thus, these systems are particularly suited for routine clinical laboratory use in diagnostic tests. Detection of pathogens causing infectious diseases and gene sequences associated with cancers or genetic diseases are among the most important uses of such tests. Reverse transcriptases are also employed as an initial step in some protocols when the polymerase chain reaction (PCR) is used to amplify an RNA target. See Malek et al., U.S. Patent No. 5,130,238 (1992); and Mocharla et al., Gene 99:271-275 (1990). In such "RT-PCR" procedures, the reverse transcriptase is used to make an initial complementary DNA (cDNA) copy of the RNA target, which is then amplified by successive rounds of DNA replication.

The retroviral reverse transcriptases have three enzymatic activities: a RNA-directed DNA polymerase activity, a DNA-directed DNA polymerase activity, and an RNase H activity. See Verma, The Reverse Transcriptase, Biochim. Biophys. Acta 473: 1-38 (1977). The latter activity specifically degrades RNA contained in an RNA:DNA duplex. Degradation of the RNA strand of RNA:DNA intermediates by RNase H is an important component of some transcription-based amplification systems and is to be distinguished from unwanted degradation due to contaminating nucleases, which interferes with amplification.

A disadvantage of the transcription-based amplification systems is their sensitivity to even trace amounts of nucleases. Since a number of important diseases may yield samples containing very few target nucleic acid molecules, the detection of small amounts of the target is often crucial for an accurate and timely diagnosis. Indeed, the value of target amplification methods is most important when the number of target molecules is low. At low input levels of the target nucleic acids, unwanted degradation of RNA targets or RNA or DNA reaction intermediates can lead to amplification failures and consequent misdiagnosis. Ribonuclease contamination is also a problem in RT-PCR reactions, since loss of the RNA target can lead to amplification failure.

Ribonucleases are relatively ubiquitous, and, in particular, are found in high concentrations in a variety of biological materials, including preparations of retroviruses and in cells commonly used to express recombinant proteins. Ribonucleases frequently contaminate reverse transcriptase preparations from a variety of sources and have been reported to interfere with synthesis of cDNAs, preparation of probes, and other uses besides target amplification alone. Often, an RNase inhibitor is included in the reaction to minimize the deleterious effects of this contamination. See e.g., Maniatis et al., Molecular Cloning: A Laboratory Manual 8.11-8.13 (2d ed. Cold Spring Harbor Laboratory Press 1989), hereby incorporated by reference herein.

However, a number of substances commonly used to inhibit or inactivate RNases, including detergents, chaotropes, organics, metals, proteases and metals are inappropriate for use in target amplification systems since they will inhibit the enzymes used for amplification as well. RNase-inhibiting proteins such as human placental RNase inhibitor, Blackburn et al., J. Biol. Chem. 252: 5904 (1977) or rat liver RNase inhibitor, Gribnau et al., Arch. Biochem. Biophys. 130: 48-52 (1969), may be unstable, are expensive, and can

contribute additional interfering substances such as nucleic acids and RNases that are not inhibited by the inhibitor.

In addition to nucleases, traces of other enzymes, nucleic acids, and certain buffer salts may interfere with amplification reactions. While these substances are merely undesirable for many uses of reverse transcriptase, because of the nature of the amplification reaction it is critical that the enzyme preparation contain as low an amount of them as possible.

Isolation and purification of reverse transcriptase from various sources have been reported. In cases where the enzyme is isolated directly from virus particles, cells, or tissues, the cost is too high for widespread commercial use in diagnostic tests. See e.g., Kacian *et al.*, *Biochim. Biophys. Acta* 46: 365-83 (1971); Yang *et al.*, *Biochem. Biophys. Res. Comm.* 47: 505-11 (1972); Gerard, *et al.*, *J. Virol.* 15: 785-97 (1975); Liu *et al.*, *Arch. Virol.* 55 187-200 (1977); Kato *et al.*, *J. Virol. Methods* 9: 325-39 (1984); Luke, *et al.*, *Biochemistry* 29: 1764-69 (1990); Le Grice *et al.*, *J. Virol.* 65: 7004-07 (1991). Additionally, these methods have not assured removal of substances that are significant inhibitors or contaminants that interfere with the use of reverse transcriptase for target amplification reactions. Another important consideration in the use of reverse transcriptases for a variety of purposes is the RNase H activity associated with the enzyme. The amount of RNase H activity and the way in which the RNase H activities work in coordination with the RNA- and DNA-dependent reverse transcriptase activities are important features affecting the utility of the enzyme for various purposes including transcription-based amplification systems. Too much or too little activity, the wrong kind of activity (such as non-specific RNases), or activities poorly coordinated with DNA synthesis can all lead to reduced performance in a particular application. Proper balance of the synthetic and degradative activities must be maintained; this is not only a function of the particular reverse transcriptase enzyme used, but also is dependent on the ability of the purification protocol to remove the RNA and/or DNA degrading activities.

The cloning and expression of reverse transcriptases in bacterial hosts has been previously reported. Attempts to clone and express reverse transcriptase from avian myeloblastosis virus (AMV-RT) did not lead to production of significant amounts of the purified enzyme. This is apparently due to the fact that the AMV-RT consists of two polypeptide chains, the  $\alpha$  and  $\beta$  chains, which must form a dimeric structure and undergo specific post-translational modifications in order to produce fully active enzyme. These same modifications do not occur when the gene is expressed in *E. coli*.

By contrast, the avian viral RTs, many reverse transcriptases derived from mammalian viruses consist of only one polypeptide chain; efforts to clone and express these enzymes have been more successful. In particular, Goff *et al.*, U.S. Patent 4,943,531 (1990) and Kotewicz *et al.*, U.S. Patent No. 5,017,492 have described methods for the purification of reverse transcriptase derived from Moloney Murine Leukemia Virus (MMLV-RT) and expressed in *E. coli*, which methods form the basis for the majority of commercial reverse transcriptase preparations.

Many commercial preparations of reverse transcriptase have been found unsuitable for use in target amplification and for other purposes due to nucleic contamination. See Sambrook, *supra*, previously incorporated by reference herein; Ryskov *et al.*, *Mol. Biol. Rep.* 8: 213-16 (1982). Other problems with commercial preparations of MMLV-RT may be related to an altered coordination between the DNA synthesis and RNase H activities of the purified enzyme, reduced ability to bind and initiate synthesis at primer sites or to read through regions of tight secondary structure, or alternately may be due to DNase and other protein contamination. See Agronovsky, A.A., *Anal. Biochem.* 203: 163-65 (1992). Additionally, commercial preparations made using the previously available methods for purification show significant lot-to-lot variability.

Moreover, due in part to the lengthy and labor-intensive purifications employed, the expense of the reagents and equipment employed for scale-up and the low yields of enzyme, the cost of such enzymes is prohibitive for their widespread commercial application in target amplification systems.

It is therefore an object of the present invention to provide an improved form of reverse transcriptase having the correct balance of DNA synthetic activities and RNase H digestive activities, thereby being particularly suited for use in nucleic acid amplification methods.

It is another object of the present invention to provide a convenient source of reverse transcriptase containing low levels of contaminants, such as undesired RNases, that interfere with transcription-based amplification reactions by cloning and expressing a gene coding for an MMLV-RT enzyme having these properties in an *E. coli* host.

It is yet another object of the present invention to reduce the RNase activity associated with the enzyme prior to and following purification by cloning and expressing the MMLV-RT gene in a ribonuclease-deficient strain of *E. coli*.

It is another object of the present invention to develop a simple purification scheme for the isolation of the enzyme.

It is a further object of the present invention to provide methods for the purification of the enzyme that achieve high levels of purity of RT at a low cost.

5

#### Summary of the Invention

The present invention features an expression vector or plasmid containing a cloned version of the gene for MMLV-RT which, when used to transform a suitable host cell such as *E. coli*, leads to the expression of the gene and the generation of a gene product having the DNA- and RNA-directed DNA polymerase activities and RNase H activity associated with retroviral reverse transcriptases.

The present invention also features a plasmid containing a gene for MMLV-RT inserted into a host cell which has a reduced level of ribonuclease activity as compared to wild-type strains.

The present invention also includes methods for the purification of the resulting enzyme from the host cells, such methods comprising suitable growth media, fermentation conditions, harvesting and storage of the cells, cell lysis and chromatography.

The present invention also features the enzyme produced by the expression vectors, host cells, and purification procedures of the present invention. The enzyme is highly-purified and suitable for use in nucleic acid amplification and other genetic engineering procedures.

Finally, the present invention features the use of the enzyme produced by the methods described herein for the synthesis of complementary DNA for a variety of purposes, notably in transcription-based amplification and RT-PCR reactions.

25

#### Brief Description of the Drawings

26

Figure 1: Construction of plasmid pUC18N

Figure 2: Oligonucleotides used to construct plasmid pUC18N.

Figure 3: Alignments of the ribosome binding sites.

Figure 4: Oligonucleotides used to modify the ribosome binding site and spacer region.

30

Figure 5: Construction of plasmid pUC18 MMLV Sst-Hind.

Figure 6: Construction of plasmid pUC18 MMLV III Tailed.

Figure 7: Oligonucleotides used to construct plasmid pUC18 MMLV Gly Tet(-).

Figure 8: Construction of plasmids pUC18N MMLV Gly and pUC18N MMLV Gly Tet(-).

35

Figure 9: Construction of plasmids pUC18N SD8D MMLV Gly and pUC18N SD8D MMLV Gly Tet(-).

Figure 10: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gel photograph of P-11 and Sephadryl S-200 fractions of purified MMLV-RT.

#### Detailed Description of the Invention

##### 40 Definitions

As used herein the following terms have the indicated meanings unless expressly indicated otherwise.

By "selectable marker gene" is meant a DNA fragment encoding a gene which, when carried and expressed by a host cell, is capable of conferring a growth advantage to that host cell as compared to cells not containing the selectable marker gene when both are grown in a culture media of a given composition. For example, the gene encoding  $\beta$ -lactamase will confer resistance to ampicillin on host cells containing this gene, whereas cells not containing the gene will be sensitive to ampicillin, thus only cells expressing the gene for  $\beta$ -lactamase will grow in media containing ampicillin. Similarly, cells unable to catabolize an essential amino acid will not grow in media not containing that amino acid, whereas cells containing a gene allowing the cell to make the essential amino acid will grow in the same media.

A selectable marker gene may be covalently linked, for example in a plasmid or expression vector, to one or more other gene or genetic element as a means of identifying cells containing both the selectable gene and the "silent" gene(s) and/or genetic element(s).

By a "purified" nucleic acid or protein is meant a nucleic acid or protein subjected to at least one step which removes cellular components such as carbohydrates, lipids, unwanted nucleic acids, or unwanted proteins from the indicated nucleic acid or protein.

By "upstream" is meant to the 5' side of a given locus on a nucleic acid strand, or in the case of a double stranded nucleic acid molecule, to the 5' side of a particular locus with respect to the direction of

gene transcription in that region of the nucleic acid molecule.

By "downstream" is meant to the 3' side of a given locus on a nucleic acid strand, or in the case of a double stranded nucleic acid molecule, to the 3' side of a particular locus with respect to the direction of gene transcription in that region of the nucleic acid molecule.

5 By " $T_m$ " is meant the temperature at which 50% of a population of a double-stranded nucleic acid molecules, or nucleic acid molecules having a double-stranded region, become single-stranded or thermally denatured.

By "recombinant" is meant that a nucleic acid molecule or protein is at least partially the result of in vitro biochemical techniques. A "recombinant DNA molecule" is thus a non-naturally occurring molecule.

10 Such recombinant molecules include, but are not limited to molecules which comprise restriction endonuclease fragments, in vitro nucleic acid ligation products, in vitro exonuclease fragments, and expression vectors comprising heterologous genetic elements such as one or more of the following: promoters, repressor genes, selectable marker genes, temperature-sensitive DNA replication elements, structural genes, and the like.

15 "Recombinant" proteins or enzymes are those not found in nature. These include purified protein preparations and proteins produced from recombinant DNA molecules. The latter proteins are usually expressed in a heterologous host cell, i.e., one not native to the protein or enzyme in question. However, the gene encoding a recombinant protein may reside on an expression vector contained within a host cell of the same species as the organism from which the protein in question was derived.

20 By "truncated" is meant a smaller version of the gene or protein in question; with respect to the primary nucleotide or amino acid sequence, a truncated form of a reference nucleic acid or protein is one that lacks one or more nucleotides or amino acids as compared to the reference molecule.

By "substantial sequence homology" is meant that a first nucleic acid or protein molecule has a recognizable non-random similarity to a second reference nucleic acid or protein over at least about 89% of 25 its nucleotide or amino acid sequence respectively.

By a nucleic acid or protein "domain" is meant at least one definite region of contiguous nucleotide or amino acid residues.

By "origin of replication" is meant a specific region of DNA at which primer production and initiation of DNA polymerase activity begins. In this specification, the term is used to mean a nucleic acid element 30 present on a DNA expression vector that allows the expression vector to increase in copy number within a given host cell.

By "promoter" is meant a genetic element comprising a specific region of DNA at which an RNA polymerase enzyme can bind and begin transcription of a DNA template, thus providing the first step of translating the genetic information contained in the sequence of a nucleic acid into the production of a 35 protein of an amino acid sequence corresponding to that nucleic acid sequence.

By "expression", "gene expression" or "protein expression" is meant the production of protein from information contained within a gene by a host organism.

By "transformation" is meant a biochemical method of inducing a host cell to internalize a nucleic acid molecule. Such nucleic acid molecules are usually genetic elements comprising at least an origin of 40 replication, a selectable marker gene, and a promoter for expression of the selectable marker gene within the host cell.

By "heterologous" is meant not of the same species. Thus, an enzyme expressed in a heterologous host cell is produced in a host cell of a different species than the one from which the enzyme was originally derived.

45 By "gene" is meant a nucleic acid region having a nucleotide sequence that encodes an expressible protein or polypeptide. A gene may comprise one or more "coding sequences" containing codons that correspond to amino acid residues of the expressed protein; the gene may also comprise, but need not comprise, one or more "non-coding" nucleotide sequence regions that do not contain codons corresponding to amino acid residues of the expressed protein.

50 All of the biochemical techniques, used for construction and evaluation of the MMLV-RT expression vectors including, but not limited to, restriction digestion protocols, gel electrophoresis, Southern blot, and DNA modification reactions, are known to those of ordinary skill in the art and are described in Sambrook et al., *supra*, which was previously incorporated by reference herein. In addition, many of these techniques are described in the co-pending application of Riggs, et al., entitled Purified DNA Polymerase from *Bacillus stearothermophilus* which enjoys common ownership with and was filed the same day as this application.

55 This application is hereby incorporated by reference herein.

I. Construction of the Cloning Vectora. Plasmid pUC18N

5 Plasmid pUC18 (Life Technologies, Inc., Bethesda, MD) was used as the parent vector. Clones were screened by restriction mapping techniques on agarose gels; such techniques are well known in the art. An Nco I restriction site was introduced between the lac Z ribosome binding site and the Eco RI restriction site of pUC18 by making a substitution of two nucleotide bases, as shown in Figure 1. The mutations were 10 introduced using the two synthetic oligonucleotides shown in Figure 2 as oligonucleotides 1 and 2 (SEQ ID NO: 1 and 2). As shown, the oligonucleotides overlap by 30 base pairs at their 3' ends. The oligonucleotides were allowed to hybridize, filled in using the Klenow fragment of *E. coli* DNA polymerase I, and digested with Pvu II and Eco RI. Plasmid pUC18 was digested with Eco RI and partially digested with Pvu II to yield two DNA fragments: a larger fragment including the intact ampicillin resistance gene (Amp), the origin of replication (Ori), and part of the lac Z gene. The smaller Eco RI-Pvu II fragment consisted of 15 the portion of the lac Z gene corresponding to positions 450 to 628 of the pUC18 map. The synthetic Eco RI-Pvu II fragment was inserted into the larger vector fragment, ligated and used to transform *E. coli* strain JM 109. Clones containing properly constructed vectors produced a blue color using an X-gal substrate (5-bromo-4-chloro-3-indolyl-β-D-galactoside) as a substrate, indicating that the lac Z gene had been properly reconstructed. These results were further verified by restriction mapping. This vector was named pUC18N. 20 (See Figure 1).

b. Construction of Plasmids Containing the Reverse Transcriptase Gene.

The intact MMLV gene was isolated as an Sst I-Hind III fragment from the pMMLV-L clone described in 25 Miller and Verma, J. Virol. 49:214-222 (1984). This fragment contained the nucleotide sequence corresponding to the region from MMLV position 2558 (Sst I site) to position 4894 (Hind III site) and contained the entire RT gene between 40 extra upstream bases and 284 extra downstream bases. Plasmid vector pUC18 was digested with Sst I and Hind III, and the vector and RT gene were ligated together and used to transform competent *E. coli* DH5αf' cells. The resulting plasmid was named pUC18 MMLV Sst-Hind (Figure 30 5). This plasmid was then digested with Eco RI and Bgl I, yielding a 2013 bp fragment of the MMLV-RT gene lacking the terminal 3' sequences of the RT gene. The RT gene fragment was ligated at its Bgl I site to a double-stranded linker designed with Bgl I-Hind III overhangs (Figure 7) from two synthetic oligonucleotides 8 and 9 (SEQ ID NOS:17 and 18, respectively). The synthetic linker contained the coding 35 sequences for the carboxyl terminus of MMLV reverse transcriptase and a stop codon. Plasmid pUC18 was digested with Eco RI and Hind III, and the large vector fragment was gel purified and ligated with the reconstructed RT gene. The resulting plasmid was called pUC18 MMLV III Tailed, and contained the MMLV gene with the extra 3' sequences removed.

c. Construction of pUC18N MMLV Gly and pUC18N MMLV Tet(-).

40 The extraneous 5' sequences of the cloned RT gene were removed as follows. A 1997 bp Mam I-Hind III fragment was isolated from pUC18N MMLV III Tailed (Figure 8). This nucleic acid fragment contained the RT gene without the 5' twenty-three nucleotides of the MMLV-RT gene sequence. Two complementary 45 oligonucleotides were synthesized and hybridized to recreate the 5' portion of the RT gene (but with initiation codon, as shown below).

Oligonucleotide #3 (SEQ ID NO:3) CATGGGTCTG AACATCGAAG  
 50 ATGA  
 Oligonucleotide #4 (SEQ ID NO:4) TCATCTTCGA TGTTCAGACC  
 5'-CATGGGTCTGAACATCGAAGATGA-3'  
 3' -CCAGACTTGTAGCTTCTACT-5'

55 Plasmid pUC18N was digested with Nco I and Hind III, and the smaller of the two resulting fragments was removed. The hybridized oligonucleotides (SEQ ID NO:3 and 4) were ligated to the larger pUC18N

fragment at the Nco I site, and the 1992 bp MMLV-RT Mam I-Hind III gene was then inserted as well, yielding the expression vector pUC18N MMLV Gly (Figure 8). The Tet gene from pUC18 Tet(+) constructed as described below, was inserted at the Aat II site, and the resulting plasmid was called pUC18N MMLV Gly Tet(-). The minus sign refers to the orientation of the Tet gene within the vector.

- 5 The cloned MMLV-RT of the present invention differs from the native enzyme in two respects. First, the codon encoding the threonine residue which occupies position 1 of the native enzyme (the second codon of the RT gene) has been replaced with a glycine codon in the cloned RT of the present invention; secondly, the codons for the leucine, asparagine and isoleucine residues occupying amino acid positions 2, 3 and 4 of the mature native protein sequence were replaced with codons more preferred by *E. coli*. The CTA codon 10 coding for leucine was replaced with degenerate codon CTG; the AAT codon coding for asparagine was replaced with degenerate codon AAC, and the ATA codon coding was replaced with degenerate codon ATC. (See Wada, K. et al., *Nucl. Acids Res.* 19(suppl.):1981-1986 (1991)).

#### d. Construction of Plasmid pUC18N SD9D

- 15 In order to optimize the expression of cloned MMLV-RT, the lac Z ribosome binding site (RBS) of pUC18N was modified to contain 9 bases complementary to *E. coli* 16S rRNA rather than on the 4 such bases present in the pUC18 parent vector. At the same time, plasmids were constructed having spacer regions separating the RBS and the ATG initiation codon by either 7, 8, or 9 base pairs, as shown for one of the strands in Figure 3 as SD7, SD8 and SD9. Common elements in the design of these spacer sequences 20 were 1) adenosine (A) in the third position 5' to the ATG initiation codon, 2) no guanine (G) or cytosine (C) in the spacer region except in the Nco I site, and 3) an 5'-RRTTTR-3' sequence spanning the RBS and the spacer, where T is thymine and R is a purine nucleotide (adenine or guanine). These common elements 25 for heterologous gene expression were suggested in Jay et al., *Proc. Natl. Acad. Sci. USA* 78: 5543-48 (1981) and Jespers et al., *Protein Engineering* 4: 485-92 (1991).

The oligonucleotides used to introduce these modifications are shown in Figure 4. Oligonucleotides 5, 6 and 7 (SEQ ID NO:5, 6 and 7) were each used in conjunction with oligonucleotide 1, shown in Figure 2. The nucleotide 4 bases on the 5' side of the ATG start codon of oligonucleotide 6 and the 4 and 5 bases on the 5' side by the ATG start codon in oligonucleotide 7 were synthesized with a mixture of A and T since neither was theoretically preferred. See Jespers, et al., supra. As in the construction of pUC18N, a 30 base pair region of complementarity existed between oligonucleotide 1 and each of oligonucleotides 3, 4 and 5. As before, each pair of oligonucleotides was allowed to hybridize, was filled in using the Klenow fragment of *E. coli* DNA polymerase I, digested with Pvu II and Eco RI and inserted into the same large pUC18 Pvu II-Eco RI fragment used in constructing pUC18N. The MMLV-RT gene was then cloned into this vector as a 35 Nco I-Hind III fragment as described below.

These constructs were evaluated by measuring the levels of MMLV-RT expression. The cells containing the plasmid with the 9-base spacer (SD9; oligonucleotide 7) displayed the highest level of reverse transcriptase expression. The plasmid was isolated and sequenced; both of the degenerate nucleotides 4 and 5 bases on the 5' side of the ATG start codon were found to be adenosine (A) residues. The expression 40 vector was named pUC18N SD9D.

#### e. Insertion of the Tetracycline Resistance Gene

The ampicillin resistance ( $\beta$ -lactamase) gene of pUC18 was used as a genetic selection marker in the 45 early vector constructions. However, owing to the fact that  $\beta$ -lactamase acts to destroy the antibiotic relatively quickly, there may be a sizable plasmid-minus revertant population in a culture in which ampicillin is the sole selective criterion.

In order to tightly regulate the cell population in the cultures, the vector was modified to contain a tetracycline resistance gene. Because tetracycline acts to block cellular uptake of the antibiotic rather than 50 inactivating it, the culture should be more stable in the presence of tetracycline than with ampicillin.

The tetracycline resistance gene was isolated from pBR322 as a 1427 bp Eco RI-Ava I fragment. The single strand overhangs were filled in using the Klenow fragment of *E. coli* DNA polymerase I, yielding a blunt-ended fragment. Aat II linkers were ligated to the tetracycline resistance gene fragment, and digested with Aat II. Plasmid pUC18 was digested with Aat II, and the linearized vector was ligated to the Aat II 55 fragment containing the tetracycline resistance gene. The ligation mixture was used to transform competent *E. coli* JM109 cells, and the transformants were selected by tetracycline resistance. The structure of the plasmid was verified by restriction mapping. Clones were selected having the tetracycline resistance gene inserted in both orientations; the plasmids were named pUC Tet(+) and pUC Tet(-).

The two plasmids were used as a supply of the tetracycline resistance gene (*Tet*) for insertions into plasmids containing cloned MMLV reverse transcriptase. This approach was preferable to attempting to insert the reverse transcriptase gene (*RT*) into a vector already containing the *Tet* gene, since the *Tet* gene contains restriction sites for enzymes used in the reverse transcriptase cloning, while the *RT* gene contains no *Aat* II sites.

**f. Construction of pUC18N SD9D MMLV Gly and pUC18N SD9D MMLV Gly Tet(-).**

- The intact, modified reverse transcriptase gene from pUC18N MMLV Gly Tet(-) was isolated as a 2018 bp Nco I-Hind III fragment and ligated with vector pUC18N SD90, from which the Nco I-Hind III polylinker region had been removed. The resulting plasmid, called pUC18N SD90 MMLV Gly, contained the MMLV-RT gene modified in the three ways described above in addition to having the improved ribosome binding site and spacer region, as described above. This plasmid was cleaved at its unique *Aat* II site, and the *Aat* II Tet gene fragment from pUC18 Tet(+) was inserted into the vector and ligated. Plasmids containing the Tet gene insert were isolated in both possible orientations, and the level of RT expression was tested for clones containing each plasmid. The clone having the Tet gene in the (-) orientation (with the coding strand in the same orientation as MMLV-RT) was found to produce higher levels of RT than the clone having the Tet gene in the opposing orientation and was therefore chosen as the preferred clone.

## 2.2 Selection of the Host Cell Strain

The following *E. coli* strains were tested for expression and purification of MMLV-RT: JM109, DH5 $\alpha$ , XL1 Blue (Stratagene, San Diego, California), JM105, ER 1458, NM 522, Inv $\alpha$  (Invitrogen, San Diego, California), TOPP™ strains 1-6 (Stratagene), 1200, MRE 600, Q 13, and A 19. Some of these strains are mutants which are deficient in RNase I (strains 1200, MRE 600, Q 13, and A 19), while others are common laboratory strains. Some of these strains contain the lac $I^R$  repressor and required induction with isopropyl-thiogalactoside (IPTG). The level of RT expression of host cells containing the RT gene was estimated by visualization of the resulting protein on SDS-polyacrylamide gels and also, in most cases, by enzymatic activity assays on crude cell lysates. Of the RNase I deficient strains, *E. coli* 1200 (Yale University *E. coli* Genetic Stock Center, strain 4449) consistently showed high levels of enzyme expression using these assays; unless indicated otherwise, all experiments described herein were conducted using this strain.

### III. Growth of *E. coli* 1200 containing pUC18N SD9D MMLV Gly Tet(-)

- The fermentation culture medium (A-Z Amine media) contained the following components in a volume of 200 liters:

N-Z Amine A (Sheffield Products, Norwich, N.Y.)	2 kg
Yeast Extract (Difco)	1 kg
NaCl	1 kg
NaOH	8 g
Tetracycline (12 mg/ml in 70% ethanol)	200 ml

- The mixture was autoclaved in the fermentation vessel at 121 °C for 20 minutes, then allowed to cool. The tetracycline was added when the temperature reached 37 °C.

The inoculum of *E. coli* 1200 containing pUC18N SD9D MMLV Gly Tet(-) was prepared by inoculating 2 ml of N-Z Amine plus 12 µg/ml tetracycline (LB + Tet) with a frozen stock culture of the vector-containing strain and incubating overnight at 37 °C with shaking. The resulting 2 ml culture was then used to inoculate 20 one-liter cultures, which were again incubated overnight at 37 °C with shaking.

The 200 liter fermenter was then inoculated with 20 liters of seed culture, and the cells were allowed to grow at 37 °C until 30 minutes after the culture had reached maximum density as determined by measuring light attenuation at a wavelength of 660 nm. This generally occurs about 7.5 hours after inoculation. During incubation the culture was stirred continuously at 150 RPM for the initial 3 hours and then at 180 RPM thereafter. The vessel was sparged with air at 45 l/min. The pH of the medium was not controlled during fermentation, and rose during that time to approximately 8.2.

The culture was chilled to 20 °C, and the cells were collected by centrifugation in a Sharpsles centrifuge. The cells were not washed. The cell paste was divided into 200 g portions and frozen in liquid N<sub>2</sub>. During

freezing, the cell mass was broken into smaller pieces to ensure rapid and thorough freezing. The frozen cell paste was then stored at -70°C.

IV. Purification of MMLV Reverse Transcriptase from E. coli 1200/pUC18N SD9D MMLV Gly Tet(-)

5

1. Assay of Reverse Transcriptase Activity and Protein Concentration.

Methods for assaying reverse transcriptase activity are known in the art. For the work described here, the dT<sub>n</sub>A assay described by Kacian was used (Kacian, Methods for Assaying Reverse Transcriptase, in Methods in Virology (Academic Press 1977), herein incorporated by reference as part of this disclosure). One unit of reverse transcriptase activity converts 1 nmole of dTTP to acid-precipitable form in 10 minutes under the conditions described therein.

15

2. Cell Lysis

15

Eleven hundred grams of frozen cell pastes were broken into pieces and suspended in 3.3 liters of Lysis Buffer (25 mM Tris-HCl (pH 7.5), 10 mM ethylenediamine tetracetic acid (EDTA), 10% (v/v) glycerol, 5 mM dithiothreitol (DTT), 1% (v/v) Triton X-100, 10 mM NaCl, 1 mM phenylmethylsulfonyl fluoride (PMSF)) by stirring at 4°C. The cells were then lysed by 2 passes through an APV Gaulin 15MR homogenizer at a continuous pressure of 8,000 psi. The receiving vessel was kept in an ice water bath, and the initial homogenate was allowed to chill for 30 minutes prior to the second pass. The lysate was then cleared by centrifugation at 4,500 x g for 1 hour at 4°C, and the pellet was discarded. The clarified lysate was either used immediately or stored frozen at -70°C and brought to 4°C before use.

25

3. Phosphocellulose Column Chromatography

30

Phosphocellulose (Whatman P11, 100 g) was treated with 2.5 liters of 0.5 N NaOH, followed by 2.5 liters of 0.5 N HCl, as recommended by the manufacturer. After a final water wash, the phosphocellulose was suspended in 1.0 l of 1.0 M Tris-HCl (pH 7.5), allowed to stand for 5-10 minutes, and transferred to a Buchner funnel. The buffer was removed by vacuum filtration, and the phosphocellulose was washed with 1.0 M Tris-HCl (pH 7.5) until the pH of the effluent matched the pH of the wash solution. The phosphocellulose was transferred to a beaker and suspended in 1.0 l of column buffer (25 mM Tris-HCl (pH 7.5), 1 mM EDTA, 10% (v/v) glycerol, 1 mM DTT, 0.1% (v/v) Triton X-100 and 1 mM PMSF) containing 0.05 M NaCl. After 5-10 minutes the buffer was removed under vacuum filtration as described above. The phosphocellulose was then suspended in 700 ml column buffer containing 0.05 M NaCl and cooled to 4°C.

35

All subsequent steps were carried out at 4°C. The chromatography was carried out using Pharmacia FPLC equipment. A Pharmacia XK 50/30 (5.0 cm x 26.0 cm) was packed with the washed and equilibrated phosphocellulose to give a bed of 500 ml. The column was then washed with 1 l of column buffer containing 0.05 M NaCl at a flow rate of 60 ml/hour. Column adapters (Pharmacia AK 50) were used to minimize the dead volume at the ends of the column. Six hundred ml of clarified cell lysate were applied to the column at a flow rate of 30 ml/hour. The column was then washed with 650 ml of column buffer containing 0.2 M NaCl at the same flow rate. Because of shrinkage of the column bed, excess buffer was removed from the space above the column bed, and the top flow adapter was readjusted to maintain contact with the bed surface.

40

The column was eluted with a 1500 ml linear salt gradient, from 0.2 M NaCl to 0.7 M NaCl in a column buffer at 30 ml/hour. The effluent was monitored for the presence of protein by its absorbance at 280 nm. Fractions of 25 ml were collected except during the elution of the protein peak, during which 15 ml fractions were collected.

45

The column fractions were analyzed using SDS-polyacrylamide gel electrophoresis (SDS-PAGE) followed by Coomassie Brilliant Blue staining. SDS-PAGE is well known in the art, and is described in Laemmli, U.K., Nature 227:680 (1970), which is hereby incorporated by reference herein. Ten microliters from each fraction were analyzed in each gel lane. A control lane contained a known amount of purified MMLV-RT. Those fractions containing a significant amount of protein migrating with an apparent molecular weight similar or identical to that of the MMLV-RT control and which contained little visible contaminating protein bands were pooled. Approximately 95% of the protein eluting within the major protein peak was able to be pooled without including a significant amount of contaminating proteins. Activity assays may also be used to locate and pool the peak MMLV-RT enzyme fractions; such procedures are known to those of ordinary skill in the art.

**4. Sephadryl S200 Gel Filtration**

The pooled phosphocellulose fractions, having a volume of 80-100 ml, were concentrated to less than 25 ml by ultrafiltration in an Amicon ultrafiltration cell using an Amicon P30 membrane at 20 psi of nitrogen.

5 Two 2.6 cm x 94 cm Pharmacia XK 26/100 columns were packed with Sephadryl S200 (Pharmacia) according to the manufacturer's directions. Column adapters were used to minimize the dead volume. Both columns were connected in series. The columns were washed with 2 l column buffer containing 0.2 M NaCl at a flow rate of 90 ml/hour. The concentrated phosphocellulose Pool (about 25 ml) was loaded onto the upstream column, and the column was developed with the same buffer at a flow rate of 90 ml/hour. Again, 10 the effluent was monitored for its absorbance at 280 nm; the initial 200 ml of effluent was collected in a single pool, and 4 ml fractions were collected during the elution of the protein peak. The MMLV-RT eluted when approximately 280-300 ml of buffer had been applied to the columns.

The fractions were again analyzed using SDS-PAGE as above. Three microliters from each fraction in the peak region were run in each gel lane; as before, a control lane contained purified MMLV-RT of a known mass. Those fractions containing a significant amount of protein migrating with the purified MMLV-RT and which contained little visible contaminating protein were pooled. Preferably, those fractions containing predominant bands of a higher apparent molecular weight than the MMLV-RT were not included in the pool. Between 95-98% of the protein in the major S200 peak was included in the Pool. Although assays for reverse transcriptase activity may be used to locate and identify the MMLV-RT in the fractions, analysis 20 preferably includes SDS-PAGE to avoid including higher molecular weight contaminants in the pool.

The pooled S200 fractions are sufficiently concentrated for most uses. The enzyme can be stored in 50% glycerol at -20°C.

**Example 1: Expression of MMLV-RT by *E. coli* containing pUC18N MMLV Gly Tet(-) or pUC18N MMLV Gly Tet(-) with a Modified Ribosome Binding Site and Spacer Sequences of Different Lengths**

The MMLV-RT gene containing the glycine amino acid substitution in the first position was evaluated in vector pUC18N and pUC18N with the spacers and modified ribosome binding site described above. All vectors contained the Tet gene and were evaluated in *E. coli* strain 1200.

30 Fifty ml cultures of *E. coli* 1200 containing either of these two constructs were grown for 16.5 hours at 37°C with shaking. Aliquots of 0.5 ml were harvested, centrifuged for 2 minutes in a microcentrifuge, and the supernatants were discarded. The cell pellets were resuspended on 0.5 ml of a wash buffer (50 mM Tris-HCl (pH 8.0), 10 mM NaCl, 5 mM EDTA and 0.25 M sucrose) and then centrifuged as before. The cell pellets were frozen at -80°C and then resuspended in 200 µl of lysis buffer (10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 1 mM EDTA, 1% glycerol, 5 mM DTT, 0.2 mM PMSF and 100 µg/ml lysozyme) and left on ice for 20 minutes. One hundred microliters of 0.75% (v/v) Triton X-100 was added to each sample, and the mixture was frozen and thawed twice. The lysate was cleared by centrifugation, and total protein was assayed by the method of Read and Northcote (*Anal. Biochem.* 116:53-64 (1981)), the disclosure of which is incorporated by reference herein.

40 Aliquots of the lysate were assayed for reverse transcriptase activity. The level of reverse transcriptase activity in each clone was calculated in terms of units per microgram of total protein in the lysate, as well as units per ml of bacterial culture. The results shown in Table 1 indicate that the vector containing the modified ribosome binding site (RBS) and the 9 base spacer sequence expressed the highest levels of enzyme.

45

Table 1

Comparison of RT Expression in Different Plasmid Constructs			
50	Expression Vector	RT activity; U per µg total protein	RT activity; U per ml culture
	unmodified pUC18N	1.81	746
	pUC18N with 7 base spacer and improved RBS	2.25	823
55	pUC18N with 8 base spacer and improved RBS	1.72	679
	pUC18N with 9 base spacer and improved RBS	2.69	1,036

## Example 2: Comparison of Modified MMLV-RT in E. coli 1200 and JM 109 Host Strains

Plasmid pUC18N was used to create plasmids encoding MMLV-RT with glycine, alanine, or valine substitutions in the first native amino acid position. These substitutions were created using oligonucleotides similar to oligos 3 and 4, but with a codon of sequence 5'-GTT-3' or 5'-GCT-3' (coding for valine or alanine respectively) in the second position of the RT gene, following the initiation codon. The Tet gene from pUC18 Tet(+ ) was inserted into the resulting plasmids in each orientation for comparison. These plasmids were used to transform *E. coli* JM109 host cells which contain an episomal copy of the lac repressor lac I<sup>q</sup>. The transformant cells were grown overnight as in Example 1, except when the cells reached log phase growth, the lac promoter was induced by the addition of 0.5 mM (IPTG) for approximately 22 hours. Aliquots were harvested and assayed for reverse transcriptase activity as in Example 1. The results are shown in Table 2 below. As can be seen, the Gly Tet(-) construction showed the highest level of enzyme expression.

15

Table 2

Effect of Orientation of Tet gene on RT Activity		
	RT Activity; U per $\mu$ g Total Protein	RT Activity; U per ml Culture
Gly Tet (+)	0.44	177
Gly Tet (-)	1.29	472
Ala Tet (+)	0.59	229
Ala Tet (-)	0.54	243
Val Tet (+)	0.91	400
Val Tet (-)	1.03	395

In a separate experiment, the Gly Tet(-) and the Val Tet(-) constructs were evaluated in *E. coli* hosts 1200 and JM 109. The JM 109 cultures were induced as above, while the 1200 cultures were uninduced. The results shown below in Table 3 indicate that the levels of expression in both strains are comparable for the Gly substituted MMLV-RT, and higher in strain 1200 for the Val substituted plasmid.

35

Table 3

Comparison of RT Expression in Different Host Cell Strains		
	RT Activity; U per $\mu$ g Total Protein	RT Activity; U per ml Culture
1200/Gly Tet (-)	1.04	591
JM 109/Gly Tet (-)	1.05	533
1200/Val Tet (-)	1.00	516
JM 109/Val Tet (-)	0.61	357

Example 3: Growth of *E. coli* 1200/pUC18N SD9D MMLV Gly Tet(-) and Expression of MMLV-RT

50

One liter of growth medium contained 10 g of N-Z Amine A, 5 g of yeast extract, 5 g of NaCl, and 0.1 ml of 10 N NaOH. One milliliter of 12 mg/ml tetracycline in 70% ethanol was added to the cooled, autoclaved medium.

55

Two ml of medium was inoculated from a frozen stock culture of the *E. coli* transformant. This was allowed to grow overnight with shaking at 37°C. The two ml bacterial culture was used to inoculate 500 ml of medium, and this culture was grown overnight as above. The 500 ml culture was, in turn, used to inoculate 5 liters of medium in a New Brunswick BioFlo III fermenter. The culture was grown at 37°C with stirring at 350 RPM. The culture was sparged with air at 4 liters/minute during fermentation. Five to ten

milliliter samples were taken every hour for measurement of pH, optical density, protein concentration and reverse transcriptase activity. These results are shown in Table 4 below.

Table 4

Growth Kinetics of 1200/pUC18N SD9D MMLV Gly Tet(-)						
Sample	Time(hr)	pH	A600	Protein (mg/ml)	RT Activity; U/Assay	RT Activity; U/mg Protein
1	0*	6.93	0.00	0.00	0.00	0
2	0"	7.15	0.29	0.06	1.17	2920
3	1	7.14	0.22	0.18	0.91	2270
4	2	6.97	0.69	0.07	0.51	1270
5	3	6.91	0.96	0.11	0.87	2170
6	4	7.02	1.72	0.09	0.92	2300
7	5	7.40	2.16	0.13	0.60	1510
8	6	7.70	2.50	0.11	1.27	3180
9	7	7.83	2.91	0.12	1.49	3720
10	8	7.98	2.79	0.09	1.46	3660

\*Pre-Inoculation

\*\*Post-Inoculation

## Example 4: Large-Scale Purification of Cloned MMLV-RT

The enzyme was prepared as described in Example 1 above. Volumes of reagents were adjusted in proportion to the weight of the pelleted cells at the beginning of the procedure. As indicated in Table 5, high purified enzyme was recovered with a 48% yield.

Table 5

Purification Parameters: MMLV-RT Purification Scale-Up					
Fraction	Volume (ml)	Protein (mg)	Total Activity (U)	Specific Activity (U/mg)	Yield (%)
Crude Lysate	605.6	2.1 x 10 <sup>4</sup>	1.5 x 10 <sup>8</sup>	7,100	100
P-11 Pool	15.2	741	8.2 x 10 <sup>7</sup>	110,656	52
Sephacryl Pool	79	363	7.9 x 10 <sup>7</sup>	217,400	48

## Example 5: SDS-PAGE of Purified MMLV-RT from 1200/pUC18N SD9D MMLV Gly Tet(-) Clone

The progress of the purification was monitored by SDS-PAGE analysis of protein in the P-11 pool and the Sephadryl pool of Example 4 above. SDS-PAGE was conducted in a 10% reducing gel essentially as described in Laemmli, *supra*. Samples were prepared as follows. An aliquot of the P-11 pool was diluted 50-fold into a gel sample buffer (50 mM Tris-HCl (pH 6.8), 10% (v/v) glycerol, 5%  $\beta$ -mercaptoethanol (BME), 2% (w/v) SDS and 0.05% (w/v) bromophenol blue) and heated at 95 °C for five minutes. An aliquot from the Sephadryl column pool was diluted 10-fold with gel sample buffer and heated in the same way. A sample of commercially obtained MMLV-RT (USB, Cleveland, OH) was prepared identically. The latter sample was reported by the supplier to have a specific activity of 187,000 U/mg and was provided in an initial concentration of 1500 U/ $\mu$ l. Prestained molecular weight markers (Bio Rad Laboratories, San Rafael, CA) were used to estimate the molecular weights of the proteins contained in the sample pools. The apparent

molecular weight of the marker proteins were 18,500 Da (egg white lysozyme), 27,500 Da (soybean trypsin inhibitor), 32,500 Da (bovine carbonic anhydrase), 49,500 Da (chicken ovalbumin), 80,000 Da (bovine serum albumin), and 106,000 Da (phosphorylase B from rabbit muscle). The gel was loaded as shown in Table 5, and is shown in Figure 10.

5

Table 5

10

15

20

25

Order of SDS-PAGE Samples in MMLV-RT Purification		
Lane	Sample	Volume ( $\mu$ l)
3	P-11 Pool	2.0
4	P-11 Pool	4.6
5	P-11 Pool	7.0
6	P-11 Pool	10.0
7	Sephacryl Pool	6.5
8	Sephacryl Pool	5.0
9	Sephacryl Pool	4.2
10	Sephacryl Pool	3.3

#### Example 6. Contaminating Ribonuclease Activity in a Commercial Preparation of MMLV-RT

30 A 24 cm x 0.4 cm column of Sephadex G75 was equilibrated with the following buffer (1X Column Buffer): 20 mM Tris-HCl (pH 7.6), 0.1 mM EDTA, 200 mM NaCl, 1 mM dithiothreitol (DTT), 0.01% (v/v) Nonidet P-40 and 10% (v/v) glycerol.

35 RNase assays were performed by using nucleic acid hybridization to measure loss of RNA incubated with the enzyme. Details of the method can be found in Arnold, et al., (U.S. Patent No. 5,283,174) and in Nelson, et al., (U.S. Patent Application No. 08/094,577). Five ml from each enzyme sample were transferred to a test tube. Ten ml of an in vitro synthesized RNA transcript (about 1-4 fmol) in water were added, and the reactions were incubated at 37°C for 1 hour. Fifty ml of an acridinium ester-labelled DNA probe complementary to a region of the RNA transcript were added in 0.1 M lithium succinate (pH 4.7), 1.1 M lithium chloride, 2% (w/v) lithium lauryl sulphate, 20 mM EDTA, 20 mM ethylene glycol bis (beta-amino ethyl ether) N, N, N', N' tetraacetic acid (EGTA), 15 mM Aldrichitol (Aldrich Chemical Company, Milwaukee, Wisconsin), and the reaction mixture was incubated at 60°C for 20 minutes. Three hundred ml of a solution of 0.6 M sodium borate (pH 8.5), 1% (v/v) Triton X-100 were added, and the reaction mixture was incubated at 60°C for 7 minutes to destroy acridium ester present on unlabelled probe. The amount of remaining label was determined in a luminometer.

40 Similar methods for assessing RNase activity using radiolabel probes or directly measuring degradation of radiolabelled RNA by monitoring conversion from acid-precipitable to acid soluble forms are well known to those skilled in the art and may be used in the practice of the present invention. Other methods for assaying low level RNase activity are available in the scientific literature and their application to the practice of the present invention is easily appreciated by those of skill in the art.

45 Twenty five microliters of a commercial preparation of MMLV-RT (U.S. Biochemicals, Cleveland, Ohio) was mixed with 12.5  $\mu$ l of 10X Column Buffer without glycerol, 10  $\mu$ l of a 10 mg/ml solution of Blue Dextran, and 77.5  $\mu$ l water. Before use the water was treated with diethyl pyrocarbonate, as described in Sambrook et al., *supra*, to destroy contaminating RNases. The enzyme was applied to the column and eluted with Column Buffer at a flow rate of 1.8 ml/hour. Two hundred thirty  $\mu$ l fractions were collected and assayed for reverse transcriptase and RNase activities, as described above.

50 The results of two identical column runs are shown in the following table.

Table 6

Comparison of Enzyme Activities For Two Different Column Runs				
	Column 1		Column 2	
Fraction	RT Activity (RLU)	RNAse Activity (% degraded)	RT Activity (RLU)	RNAse Activity (% degraded)
1				
2	61			
3				
4	1000			
5				
6	2574			
7				
8	1751		1605	
9			20029	
10	14328	14	98216	0
11				
12	143493	0	40619	0
13			43523	
14	21570	51	9299	55
15			26650	
16	11306	0	13490	17
17			1226	
18			4713	
19			1462	
20	2583	64	1379	57
21			1263	
22	913	52	1072	0
23				
24	907	21		64
25				
26	987	64		73
27				
28	21375			
29				
30	8100			

As this Table illustrates, the commercial enzyme preparation contains significant endogenous RNAse activity. This RNAse activity is other than the RNAse H activity associated with the MMLV-RT enzyme, since it degrades single-stranded RNA. When analyzed by gel filtration chromatography, at least four peaks of non-RNAse H RNAse activity are obtained. These peaks may represent four distinct enzymes. Alternatively, they may represent aggregation of one or more protein, dissociation of such a protein into subunits, or other chromatographic artifacts. At least one of these peaks of non-RNAse H RNAse activity co-

elutes with the MMLV-RT.

Example 7: Comparison of Contaminating Ribonucleases in Partially Purified Recombinant MMLV-RT from E. Coli Host Cells JM 109 and 1200

In order to compare the amount of contaminating ribonuclease activities present in MMLV-RT-containing cell lysates after P11 column purification between host cells JM 109 and 1200 transformed with plasmid pUC18N SD9D MMLV Gly Tef(-), fractions from each column were assayed for reverse transcriptase activity using the dT:R assay described in Kacian, *Meth. Virol. supra*, and for non-RNAse H RNase activity using the assay described in the previous example.

The results obtained for each cell type are shown in the following tables:

Table 7

E. Coli Strain 1200		
Fraction	RT Activity (RLU)	RNAse activity (% degraded)
1	1007	0
5	1084	0
10	1021	0
15	3712	0
20	38359	0
25	20741	0
30	316513	0
33	346922	0
36	504196	0
39	387533	0
42	371897	0
45	472248	0
48	1199993	0
51	1529015	0
54	1126592	0
57	1034428	0
60	850009	0
63	698462	0
66	390121	0
69	177736	0
72	260049	0
76		

Table 8

E. Coli Strain JM 109		
Fraction	RT Activity (RLU)	RNAse activity (% degraded)
1	1103	0
5	1287	0
14	1287	0
15	28359	29
20	50927	75
25	29551	70
30	350732	30
35	198151	30
38	164047	54
41	149647	66
44	161963	62
47	674123	81
50	2060603	83
53	2703286	85
56	1967435	
59	1608490	90
62	782936	86
65	265569	78
68	147948	63
71	78481	38
74	44426	3
77	19964	0
81	13900	0

The data show that the enzyme prepared from JM 109 cells contained significant amounts of non-  
RNAse H ribonuclease activity throughout the P11 column profile. Significant amounts of RNAse activity  
eluted with the reverse transcriptase activity. In contrast, the reverse transcriptase purified from the E. coli  
1200 cells was free of detectable contaminating RNAse activity after the crude extract was purified by  
phosphocellulose column chromatography.

Example 8: Amplification of Mycobacterium tuberculosis Ribosomal RNA Target Sequence Using Purified  
Recombinant MMLV Reverse Transcriptase From E. coli 1200/pUC18N SD9D MMLV Gly Tet(-)

Nucleic acid amplification was performed using the procedure described in Kacian and Fultz, EPO 0  
408 295 A2, which is incorporated by reference herein and which enjoys common ownership with the  
present application. A reagent mixture was made up as follows: 768 microliters of water was given, in order,  
25  $\mu$ l 1 M Tris-HCl (pH 8.0), 50  $\mu$ l 1 M MgCl<sub>2</sub>, 44  $\mu$ l KCl, 500  $\mu$ l 40 mM rNTPs, 500  $\mu$ l 10 mM dNTPs, 9  $\mu$ l

T7 promoter-primer (84 pmoles/ $\mu$ l), and 5  $\mu$ l non-T7 primer (150 pmole/ $\mu$ l) and mixed. The volume of this mixture (Solution A) was calculated to be suitable for 50 assays. Forty  $\mu$ l of solution A was added to each reaction tube. Ten microliters of the purified target rRNA (0.05-25 fg/ $\mu$ l diluted in Template Dilution Buffer (0.2% (w/v) bovine serum albumin in 150 mM NaCl)) was added to each tube. The target rRNA had nucleic acid sequences sufficiently complementary to the primer and the promoter-primer to allow hybridization to occur under stringent hybridization conditions. Preparation of rRNA is known to those of skill in the art. Two hundred microliters of silicone oil was layered onto the surface of each reaction mixture, and the reaction tubes were heated at 95 °C for 15 minutes in a heating block. The reaction tubes were then transferred into a 42 °C water bath and allowed to cool for 5 minutes.

An enzyme mixture was prepared by transferring 46.8  $\mu$ l Dilution Buffer to a tube and adding 1.1  $\mu$ l (900 U) MMLV-RT and 2  $\mu$ l (400 U) T7 RNA polymerase. This mixture was then added to each tube. The reactions were then incubated at 42 °C for two hours.

The amount of amplified RNA generated was then measured using an acridinium ester-labeled DNA probe directed to the target sequence as described in Arnold et al., PCT WO89/02476 and Arnold et al., Clin. Chem. 35:1588-1594 (1989) the former of which enjoys common ownership with the present invention, and both of which are incorporated by reference herein. All reactions were run in quadruplicate except for the negative control, which was run in duplicate. The results shown in Table 9 below indicate that saturating levels of the amplified target sequence are obtained with as little as 2.5 fg of input template RNA at the beginning of the experiment.

20

Table 9

Sensitivity of Enzyme Preparation in an Amplification Reaction		
	Amount of Template RNA Added (fg)	Signal (RLU)
25	250	2841164
		2802308
		2828732
		2828837
30	25	2801357
		2968585
		2748909
		2723562
35	2.5	2761901
		2809799
		2932942
		2906826
40	0	2246
		2443

Example 9: Synthesis of cDNA using Purified Recombinant MMLV-RT from E. coli 1200/pUC18N SD9D  
50 MMLV Gly Tet(-)

The ability of the recombinant purified MMLV-RT to synthesize cDNA was compared to that of a commercially available reverse transcriptase preparation (U.S. Biochemicals) in an RNA sequencing reaction.

55 TTE buffer was prepared by mixing 20 ml 1 M Tris-HCl (pH 7.5), 0.4 mM EDTA (pH 8.0) and 281.7  $\mu$ l triethylamine. Primers had the following sequences:

SEQ ID NO:8 5'-TACCTTGTTACGACTCACCCCA-3'

SEQ ID NO:9 5'-CTTAGATGCTTCAGC-3'

The primer were labeled with  $^{32}\text{P}$  at their 5' ends using polynucleotide kinase; procedures for end-labeling nucleic acids are generally known in the art. After being end labelled, the primers were purified by chromatography on Nensorb™ columns (New England Nuclear) according to the manufacturer's specifications, followed by ethanol precipitation.

- 5 Reactions were carried out using either purified recombinant MMLV-RT from *E. coli* 1200/pUC18N SD9D MMLV Gly Tet(-) or reverse transcriptase purchased from a commercial vendor.

Reaction mixtures contained the following reagents in 100  $\mu\text{l}$  final volume: Ten microliters of GPE Buffer (500 mM Tris-HCl (pH 7.6), 175 mM MgCl<sub>2</sub>, 250 mM KCl, 20 mM spermidine), 8  $\mu\text{l}$  of stock rNTPs (25 mM rCTP and rUTP; 65 mM rATP and rGTP), 4  $\mu\text{l}$  of stock dXTPs (10 mM), 0.5  $\mu\text{l}$  1 M DTT, 20 pmoles  $^{32}\text{P}$ -labelled primer, 20 pmole unlabelled primer, 20 pmoles purified *E. coli* rRNA, 600 U reverse transcriptase. Reactions were established by mixing all components without the reverse transcriptase, then heating the mixture at 95 °C for 5 minutes to denature the template RNA secondary structure. Reactions were then placed at 60 °C for 30 minutes to allow the primers to anneal to the rRNA target. The reaction mixture was cooled to room temperature, and the reverse transcriptase was added. DNA synthesis was carried out at 42 °C for 60 minutes. Reactions were analyzed on 7% polyacrylamide gels essentially as described in Williams et al., BioTechniques 4: 138-147 (1986).

Both enzymes were found to synthesize cDNA from the RNA template with equal efficiency as judged from the gel electropherograms.

- 20 Example 10: Reverse Transcriptase-Mediated PCR using Recombinant MMLV-RT from *E. coli* 1200/pUC18N SD9D MMLV Gly Tet(-)

All PCR reactions were run in a Perkin Elmer-Cetus Model 9600 DNA thermal cycler. The thermal cycler was programmed to incubate the reaction in the following manner and sequence:

- 25 94 °C for 3 minutes;  
35 cycles between 51 °C for 30 seconds, 72 °C for 2 minutes, and 94 °C for 1 minute;  
72 °C for 5 minutes;  
4 °C overnight.

Two separate preparations of MMLV-RT were used for this experiment, as well as a lot from the same commercial vendor as above. Different amounts of RT were tested, but 50 U of the enzyme was found to be optimum for all enzyme preparations used. The reagents used in the experiment were as follows: 5X RT Buffer (50 mM Tris-HCl (pH 8.3), 75 mM KCl, 3 mM MgCl<sub>2</sub>, 5 mM DTT); 10X PCR Buffer (Perkin Elmer)-(100 mM Tris-HCl (pH 8.3), 500 mM KCl, 15 mM MgCl<sub>2</sub>, 0.1% gelatin); RT Premix (for each reaction) (4  $\mu\text{l}$  5X RT Buffer, 0.8  $\mu\text{l}$  of a 25 mM solution of each dNTP, 50 units RT, 100 moles (-) sense primer, water in a total volume of 20  $\mu\text{l}$ ); PCR Premix (for each reaction) (8  $\mu\text{l}$  10X PCR Buffer, 100 pmoles (+) sense primer, 2.5 units Taq DNA Polymerase, and water to a total volume of 80  $\mu\text{l}$ ). Probes were stored in 10 mM lithium succinate buffer (pH 5.0), 0.1% lithium lauryl sulfate (LLS).

The probes and primers used for this experiment were designed to be complementary to sequences of the human papilloma virus (HPV) genome. The probes were labeled with acridinium ester as disclosed in Arnold and Nelson, PCT Patent Application No. WO89/02476, hereby incorporated by reference herein.

40 Crude preparations of unspliced template RNA were made by suspending SiHa cells (which contain HPV nucleic acid sequences integrated into their genome) at a concentration of  $1.6 \times 10^7$  cells/ml in 10 mM sodium phosphate (pH 7.8), 100 mM NaCl. The cells were heated for 15 minutes at 95 °C, cooled to room temperature, then diluted into water to the desired concentration. RNA transcripts from the E6 gene were prepared by *in vitro* transcription of DNA from a plasmid containing the HPV16 E6 gene. This plasmid was constructed by cloning a DNA fragment from the HPV clone described by Matsukura et al., J. Virol. 58:979-982 (1986) into pBluescript™ II SK (+) and (-) sense cloning vectors (Stratagene, San Diego, CA). These references are incorporated by reference herein. RNA transcripts were prepared as indicated by the manufacturer.

50 The amplification reactions were conducted as follows. Target nucleic acids were added to the MMLV-RT premix. This mixture was heated at 95 °C for 2 minutes. The primers were added and allowed to anneal to the target nucleic acids for 10 minutes at 60 °C. The reaction mixture was then cooled on ice. Reverse transcriptase was added, and the reaction was incubated at 37 °C for 30 minutes. The reaction was then heated at 95 °C for 10 minutes to inactivate the reverse transcriptase. The mixture was cooled in ice, and two drops of mineral oil were layered onto the surface of each tube.

55 Taq DNA polymerase, was diluted into the PCR Premix at the concentration indicated above. Eighty microliters of the PCR Premix was then added to each sample. The samples were placed in the thermal cycler at 95 °C, and cycling was performed as described above.

Hybridization and detection were carried out as described in Arnold and Nelson, *supra*. For each hybridization assay, 30  $\mu$ l of water was given 10  $\mu$ l of a PCR reaction mixture. The DNA was denatured at 95°C for 5 minutes. Ten microliters of diluted probe was added and mixed. The tubes were then incubated at 60°C for 15 minutes. Three hundred microliters of selection reagent was added, the tubes were mixed and incubated at 60°C for 5 minutes. The tubes were then cooled in ice, and the remaining acridinium ester label was measured in a LEADER™ luminometer (Gen-Probe Incorporated, San Diego, CA).

The results are shown in Table 10.

Table 10

Copies of Template RNA	RNA Type	Origin of Reverse Transcriptase	Average Net RLU
1 $\times$ 10 <sup>7</sup>	SiHa cell lysate	commercial (who?)	423,084
1 $\times$ 10 <sup>7</sup>	SiHa cell lysate	<u>E. coli</u> 1200/pUC18N SD9D MMLV Gly Tet(-)	445,003
1 $\times$ 10 <sup>7</sup>	E6 transcript	commercial	2,741,628
1 $\times$ 10 <sup>7</sup>	E6 transcript	<u>E. coli</u> 1200/pUC18N SD9D MMLV Gly Tet(-)	2,291,786
1 $\times$ 10 <sup>4</sup>	E6 transcript	commercial	103,501
1 $\times$ 10 <sup>4</sup>	E6 transcript	<u>E. coli</u> 1200/pUC18N SD9D MMLV Gly Tet(-)	1,395,572
1 $\times$ 10 <sup>5</sup>	E6 transcript	commercial	1,317,386
1 $\times$ 10 <sup>5</sup>	E6 transcript	<u>E. coli</u> 1200/pUC18N SD9D MMLV Gly Tet(-)	2,283,978
1 $\times$ 10 <sup>6</sup>	E6 transcript	commercial	1,661,390
1 $\times$ 10 <sup>6</sup>	E6 transcript	<u>E. coli</u> 1200/pUC18N SD9D MMLV Gly Tet(-)	2,951,045
1 $\times$ 10 <sup>7</sup>	E6 transcript	commercial	2,294,856
1 $\times$ 10 <sup>7</sup>	E6 transcript	<u>E. coli</u> 1200/pUC18N SD9D MMLV Gly Tet(-)	2,421,754

35

40

45

50

55

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- 5                   (i) APPLICANT:  
                       (A) NAME: GEN-PROBE INCORPORATED  
                       (B) STREET: 9880 CAMPUS POINT DRIVE  
                       (C) CITY: SAN DIEGO  
                       (D) STATE: CALIFORNIA  
                       (E) COUNTRY: USA  
 10                  (F) POSTAL CODE (ZIP): 92121  
                       (G) TELEPHONE: (213) 489-1600  
                       (H) TELEFAX: (213) 955-0440  
                       (I) TELEX: 67-3510
- 15                  (ii) TITLE OF INVENTION: HIGHLY-PURIFIED RECOMBINANT REVERSE  
                       TRANSCRIPTASE
- 16                  (iii) NUMBER OF SEQUENCES: 18
- 20                  (iv) COMPUTER READABLE FORM:  
                       (A) MEDIUM TYPE: Floppy disk  
                       (B) COMPUTER: IBM PC compatible  
                       (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
                       (D) SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
- 25                  (v) CURRENT APPLICATION DATA:  
                       APPLICATION NUMBER: EP 95104959.2

## 25                  (2) INFORMATION FOR SEQ ID NO: 1:

- 30                  (i) SEQUENCE CHARACTERISTICS:  
                       (A) LENGTH: 118 base pairs  
                       (B) TYPE: nucleic acid  
                       (C) STRANDEDNESS: single  
                       (D) TOPOLOGY: linear
- 35                  (ii) MOLECULE TYPE: DNA (genomic)
- 40                  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 GGCATGCAGC TGGCACGACA GGTTTCCCGA CTGGAAAGCG GGCAGTGAGC GCACCGCAAT      60  
 TAATGTGAGT TAGCTCACTC ACTCATTAGG CACCCAGGC TTTCACCTTT ATGCTTCC      118

## 45                  (2) INFORMATION FOR SEQ ID NO: 2:

- 40                  (i) SEQUENCE CHARACTERISTICS:  
                       (A) LENGTH: 111 base pairs  
                       (B) TYPE: nucleic acid  
                       (C) STRANDEDNESS: single  
                       (D) TOPOLOGY: linear
- 45                  (ii) MOLECULE TYPE: DNA (genomic)
- 50                  (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 CCGTGGGTCC GAAATGTGAA ATACGAAGGC CGAGCATACA ACACACCCTTA ACACTCGCCT      60  
 ATTTTGAAAG TGTGTCCCTTT GTCGGTACCG GTACTAATGC TTAAGCTCGA A      111

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GATGGGTCTG AACATCGAAG ATGA

24

## 15 (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

20 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

TCATCTTGA TGTTCAAGACC

20

25

## (2) INFORMATION FOR SEQ ID NO: 5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 114 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AGACTCGAAC TCGTAATCAT GGCCATGGTT TAAACCTCT TAGTGAAATT GTTATCCGCT  
CACAAATTCCA CACAACATAC GAGCCGGAAG CATAAACTGT AAGCCTGGGG TGCC

60

114

40

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 116 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

50

55

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AGACTCGAAT TCGTAAATCAT GGCCATGGTA TTAAACCTC CTTAGTGAAA TTGTTATCCG  
 60  
 5 CTCACAATTG CACACAACAT ACGAGCCGG AGCATAAAGT GTAAGCCTGG GGTGCC

116

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

AGACTCGAAT TCGTAAATCAT GGCCATGGTA TATTTAAACC TCCTTAGTGA AATTGTTATC  
 60  
 15 CGCTCACAAT TCCACACAAAC ATACGAGCCG GAAGCATAAA GTGTAAGCCT GGGGTGCC

118

20

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

30 AGGCAGCCAT CACAGAGACT CCAGACACCT CTACCCCTCCT CTAATA

46

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

40 AGCTTATTAG AGGAGGGTAG AGGTGTCTGG AGTCTCTGTG ATGGCTGCT TTC

53

45

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

50

55

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

5 ATGGGTCTGA ACATC

15

(2) INFORMATION FOR SEQ ID NO: 11:

10 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TAAGGAGGT

9

20 (2) INFORMATION FOR SEQ ID NO: 12:  
 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

25 RRTTTTRR

7

(2) INFORMATION FOR SEQ ID NO: 13:

30 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

35 TAAAAAACCC

9

45

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 6 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single

50

55

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATGGGT

6

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATGGGTCTGA ACATC

15

20

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TAAGGAGGTT TAAAAAAC

19

35

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 46 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGGCAGCCAT CACAGAGACT GGAGACACCT CTACCCCTCT CTAATA

46

45

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 53 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

55

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

5 AGCTTATTAG AGGAGGGTAG AGGTCTCTGG AGTCTCTGTG ATGGCTGCCT TTC

53

10

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

15

(A) LENGTH:

118

20

25

30

35

40

45

50

55

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
 GGCATGCAGC TGGCACGACA GGTTTCCCGA CTGGAAAGCG GGCAGTGAGC 50  
 GCAACGCAAT TAATGTGAGT TAGCTCACTC ACTCATTAGG CACCCCAGG 100  
 10 TTTACACTTC ATGCTTCC 118

(3) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 111  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

20 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
 CGCTGGGTCC GAAATGTGAA ATACGAAGGC CGAGCATACA ACACACCTTA 50  
 ACACCTGGCCT TAATTGAAAG TTGTGTCTTT GTCGGTACCG GTACTAATG 100  
 25 TTAAGCTCGA A 111

(4) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 3:  
 GATGGGTCTG AACATCGAAG ATGA 24

(5) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

40 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
 TCATCTTCGA TGTTCA 20 GACC

(6) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

45 (A) LENGTH: 114  
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 5  
 AGACTCGAAT TCGTAATCAT GGCCATGGTT TAAACCTCCT TAGTGAAATT 50  
 GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT100  
 114 AAGCCTGGGG TGCC

10 (7) INFORMATION FOR SEQ ID NO: 6:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 116  
 (B) TYPE: nucleic acid  
 15 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 6  
 AGACTCGAAT TCGTAATCAT GGCCATGGTA TTTAAACCTC CTTAGTGAAA 50  
 20 TTGTTATCCG CTCACAATTCC CACACAACAT ACGAGCCGGA AGCATAAAGT110  
 116 GTAAAGCCTGG GGTGCC

(8) INFORMATION FOR SEQ ID NO: 7:  
 25 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 118  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 30 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 7:  
 AGACTCGAAT TCGTAATCAT GGCCATGGTA TATTTAAACC TCCTTAGTGA 50  
 35 AATTGTTATC CGCTCACAAAT TCCACACAAAC ATACGAGCCG GAAGCATAAA100  
 118 GTGTAAAGCCT GGGGTGCC

(9) INFORMATION FOR SEQ ID NO: 8:  
 40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 46  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 45 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 8:  
 AGGCAGCCAT CACAGAGACT CCAGACACCT CTACCCCTCT CTAATA 46  
 AGGCAGCCAT CACAGAGACT CCAGACACCT CTACCCCTCT CTAATA

(10) INFORMATION FOR SEQ ID NO: 9:  
 50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 53

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
 AGCTTATTAG AGGAGGGTAG AGGTGTCCTGG AGTCTCTGTG ATGGCTGCCT 53  
 TTC

10 (11) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

15 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
 15 ATGGGTCTGA ACATC

20 (12) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

25 (iii) SEQUENCE DESCRIPTION: SEQ ID NO: 11:  
 9 TAAGGAGGT

(13) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

30 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 12:  
 7 RRTTTTRR

(14) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

35 (iii) SEQUENCE DESCRIPTION: SEQ ID NO: 13:  
 55

9

TTAAAAAACC

(15) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ATGGGT

(16) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ATGGGTCTGA ACATC

(17) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TAAGGGAGTT TAAAAAAC

(18) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 46  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGGCAGCCAT CACAGAGACT GGAGACACCT CTACCCCTCCT CTAATA

(19) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53  
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

5 (ii) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AGCTTATTAG AGGAGGGTAG AGGTGTCTGG AGTCTCTGTG ATGGCTGCCT 50

53

TTC

10

### Claims

1. A method for producing a polypeptide having RNA-directed and DNA-directed DNA polymerase activities comprising the steps:
  - a. constructing a plasmid comprising a gene encoding said polypeptide, at least one selectable marker gene, a promoter sequence and a replicon capable of autonomously replicating said plasmid within a suitable host cell,
  - b. inserting the plasmid into a suitable host cell deficient in the expression of RNase activity,
  - c. growing the host cells containing said plasmid in a liquid culture under conditions capable of promoting cell division and expression of the polypeptide gene,
  - d. lysing said host cells to form a cell lysate, and
  - e. purifying said polypeptide from the cell lysate.
2. The method of Claim 1 wherein the polypeptide comprises a single covalently linked amino acid chain.
3. The method of Claim 2 wherein the gene encoding said polypeptide comprises a nucleic acid sequence derived from Moloney Murine Leukemia Virus.
4. The method of Claim 3 wherein said polypeptide has a methionine residue at an amino terminus.
5. The method of Claim 4 wherein the polypeptide has a glycine as the penultimate residue at said amino terminus.
6. The method of Claim 3 wherein the nucleic acids encoding the five amino acid residues at the amino terminus of said polypeptide comprise codons to which E. coli tRNA will preferentially bind.
7. The method of Claim 3 wherein the nucleic acids encoding the five amino acid residues at the amino terminus of said polypeptide consist of the sequence 5'-ATGGGTCTGAACATC-3' (SEQ ID NO:10).
8. The method of any of claims 1 to 7 wherein said plasmid further comprises a ribosome binding site 7 to 9 nucleotides upstream from the initiation codon of said polypeptide gene.
9. The method of Claim 8 wherein said ribosome binding site comprises a nucleic acid sequence 45 complementary to a nucleic acid sequence of E. coli rRNA.
10. The method of Claim 9 wherein said ribosome binding site comprises the sequence 5'-TAAGGAGGT-3' (SEQ ID NO:11).
11. The method of any of Claims 8 to 10 wherein the ribosome binding site is 9 nucleotides upstream from the initiation codon of said polypeptide gene.
12. The method of any of Claims 8 to 11 wherein a spacer nucleic acid sequence between said ribosome binding site and said initiation codon comprises a sequence having an adenine residue at the position three nucleotides upstream from the initiation codon.
13. The method of Claim 12 wherein the ribosome binding site and spacer sequence together have a nucleic acid sequence comprising 5'-RRTTTRR-3' (SEQ ID NO:12), and said nucleic acid sequence

overlaps the ribosome binding site.

14. The method of Claim 12 or 13 wherein the spacer sequence consists of 5'-TTAAAAACCC-3' (SEQ ID NO:13).
- 5 15. The method of any of Claims 1 to 14 wherein the selectable marker gene confers resistance to at least one antibiotic to said host cell.
- 10 16. The method of Claim 15 wherein said antibiotic comprises tetracycline.
17. The method of any of Claims 1 to 16 wherein the polypeptide further is able to degrade an RNA strand in an RNA-DNA hybrid.
18. The method of any of Claims 1 to 17 wherein said host cell comprises Escherichia coli strain 1200.
- 15 19. A composition comprising the polypeptide produced by the method of any of Claims 1 to 18.
- 20 20. A method for purifying a recombinant polypeptide having RNA-directed and DNA-directed DNA polymerase activities wherein said polypeptide is expressed in an RNase-deficient bacterial host cell comprising:
  - a. lysing a plurality of said host cells containing the recombinant polypeptide thereby forming a cell lysate,
  - b. applying the cell lysate to a cation-exchange medium in the presence of a solution having a conductivity corresponding to no more than about 0.05 M NaCl thereby binding said polypeptide to the cation-exchange medium,
  - c. eluting the bound polypeptide from the cation-exchange medium by contacting said polypeptide with a salt gradient beginning with a conductivity corresponding to about 0.2 M NaCl and ending with a conductivity corresponding to about 0.7 M NaCl, and recovering the polypeptide in at least one fraction,
  - d. applying the recovered polypeptide to a gel filtration column and again recovering said polypeptide in at least one fraction.
- 30 21. The method of Claim 20 wherein the RNase-deficient host cell consists of E. coli strain 1200.
- 35 22. The method of Claim 20 or 21 wherein the cation-exchange medium comprises a phosphocellulose derivative.
23. The method of any of Claims 20 to 22 wherein the polypeptide has an apparent molecular weight of about 70,000 daltons.
- 40 24. A composition comprising a polypeptide having RNA-directed and DNA-directed DNA polymerase activities and RNase H activity, in a medium essentially free of additional nuclease activities.
25. A recombinant DNA molecule comprising:
  - a. a DNA fragment encoding a polypeptide having RNA-directed and DNA-directed DNA polymerase activities,
  - b. a DNA fragment comprising at least one selectable marker gene,
  - c. a DNA fragment comprising a promoter sequence for expressing said polypeptide gene, and
  - d. a DNA fragment comprising an origin of replication capable of promoting the autonomous replication of said recombinant DNA within a suitable host cell, said DNA fragments being so linked that the protein having RNA-directed and DNA-directed polymerase activities are expressed by said host cell.
- 45 26. The recombinant DNA of Claim 25 wherein the polypeptide comprises a single covalently linked amino acid chain.
- 50 27. The recombinant DNA of Claim 26 wherein the gene encoding said polypeptide comprises a nucleic acid sequence derived from Moloney Murine Leukemia Virus.

28. The recombinant DNA of Claim 27 wherein the gene for said polypeptide begins with the nucleotide sequence 5'-ATGGGT-3' (SEQ ID NO:14).
29. The recombinant DNA of any of Claims 25 to 28 wherein the fifteen nucleotides comprising the 5' terminus of said polypeptide gene comprise codons to which E. coli t-RNA will preferentially bind.
30. The recombinant DNA of Claim 29 wherein the nucleotides comprising the 5' terminus of said polypeptide gene consist of the sequence 5,-ATGGGTCTGAACATC-3' (SEQ ID NO:15).
31. The recombinant DNA of any of Claims 25 to 30 further comprising a ribosome binding site from seven to nine nucleotides upstream from an initiation codon of said polypeptide gene.
32. The recombinant DNA of Claim 31 wherein said ribosome binding site comprises a nucleic acid sequence sufficiently complementary to a nucleic acid sequence of E. coli r-RNA.
33. The recombinant DNA of Claim 32 wherein the ribosome binding site comprises the sequence 5'-TAAGGAGGT-3' (SEQ ID NO:11).
34. The recombinant DNA of any of Claims 25 to 33 wherein a spacer nucleic acid sequence between said ribosome binding site and said initiation codon comprises a nucleic acid sequence having an adenine residue at the position three nucleotides upstream from the initiation codon.
35. The recombinant DNA of Claim 34 wherein the ribosome binding site and the spacer sequence together have a nucleic acid sequence comprising SEQ ID NO:12 where R is adenosine or guanine and said nucleic acid sequence overlaps the ribosome binding site.
36. The recombinant DNA of Claim 34 wherein the spacer sequence consists of the nucleic acid sequence 5'-TTAAAAAACC-3' (SEQ ID NO:13).
37. The recombinant DNA of Claim 35 wherein the spacer sequence consists of the nucleic acid sequence SEQ ID NO:13.
38. The recombinant DNA of Claim 37 wherein the ribosome binding site, the spacer sequence, and the initiation codon together comprise a sequence of contiguous bases consisting of the sequence 5'-TAAGGAGGTTTAAAAAACC-3' (SEQ ID NO:16).
39. The recombinant DNA of any of Claims 25 to 38 wherein the selectable marker gene confers resistance to at least one antibiotic to said host cell.
40. The recombinant DNA of Claim 39 wherein said antibiotic comprises tetracycline.
41. The recombinant DNA of any of Claims 25 to 40 wherein said polypeptide-encoding DNA fragment further encodes a polypeptide able to degrade an RNA strand in an RNA-DNA hybrid.

45

50

55

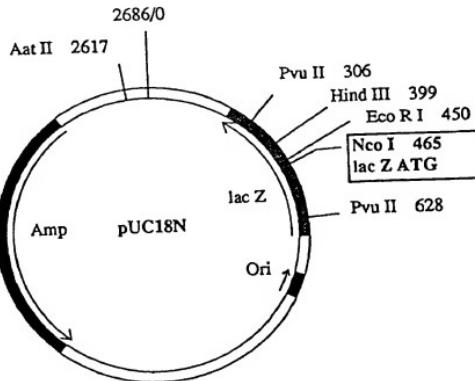


FIG. I.

## FIG. 2.

Oligo #1	5'. <u>GGCATTGCA</u> GACAGGTTCGGACTGGAAAGCGGGCAGTGAGCGAACGC Pvu II site
Oligo #1	AATTAA <u>TGTGAGTTAGCTCACTCATTAGGCACCCAGGCTTACACTTATGCTTCC-3'</u>
Oligo #2	3'. CCGTGGGTCCGAATGTGAAATA <u>CGAAGG</u>
Oligo #2	CCGAGCATACACACACCTTAACACTCGCCTATTTGAAAGTGTCTTTGTGGT ACCGGTACTAAT <u>GCTTAAGCTCGAA-5'</u> Eco RI site

DISTANCE FROM  
RBS to ATG

FIG. 3.

E coli 16S rRNA 3'-5'  
(3' end) ATT CCT CACTAGGT.....

	DISTANCE FROM RBS to ATG
pUC lac Z	c a c a c AGGA A A C A G C T A T G c a c a c AGGA A A C A G C C A T G G
pUC 18 N lac Z	c a g c c GGAGTAGA A GATG
Jespers et al "R" gene (natural)	A A G G A G G T T T A A A T A T G A T A A G G A G G T T T A A A T A T G
Jespers et al "R" gene (improved)	A T A A G G A G G T T T A A A T A T G (11-12) G C A T A A G G A G G T T T A A A G C T (4-5 bases)
Jay et al synthetic improved	C A C T A A G G A G G T T T A A A C C A T G G C A C T A A G G A G G T T T A A A C C A T G G C A C T A A G G A G G T T T A A A C C A T G G C A C T A A G G A G G T T T A A A C C A T G G
SD 7	
SD 8	
SD 9	
	BR T T T R R Aat between RBS and ATG CCATGG = Nco I site A at -3 position

FIG. 4.

AAATTGTATCCGCTACAATTCCACACAACATAAGGCCGGAAAGCAT

Oligomers #6 and 7 are identical to oligomer #5 except as indicated.

Oligomers #5, 6 and 7 were used with oligomer #1 as in the construction of PUC18N.

FIG. 5.

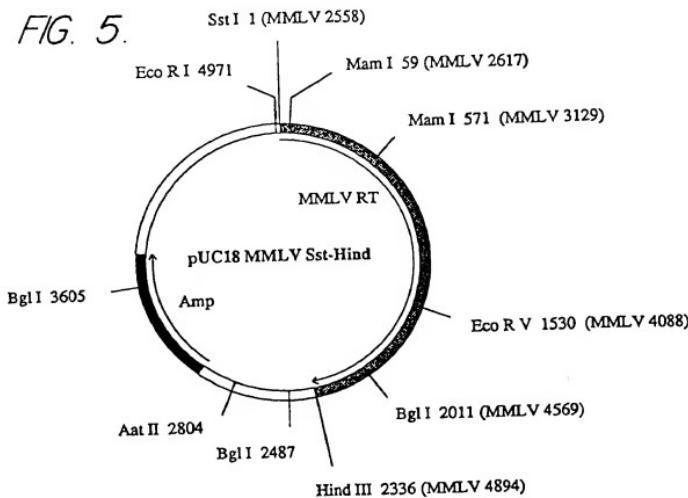


FIG. 6.

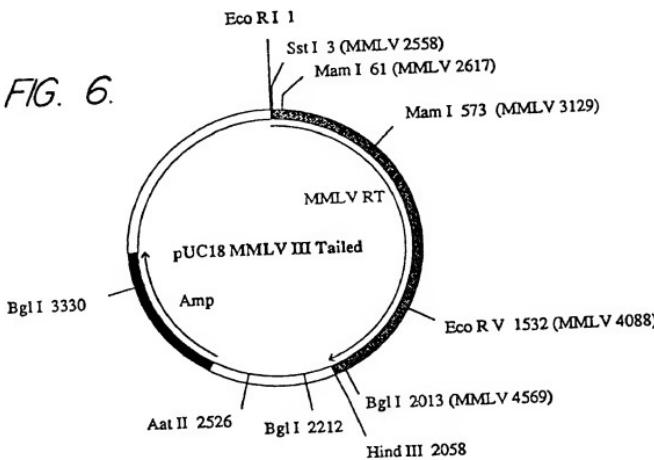


FIG. 7

5'- AGGCAGCCATCACAGAGACTCCAGACACCTTACCCCTCCTCTAAATA -3'  
Oligomer #8  
3'- CTTTCCGGTGGTAGTGTCTGAGGCTGTGGAGATGGAGATTTCGA-5'  
Oligomer #9

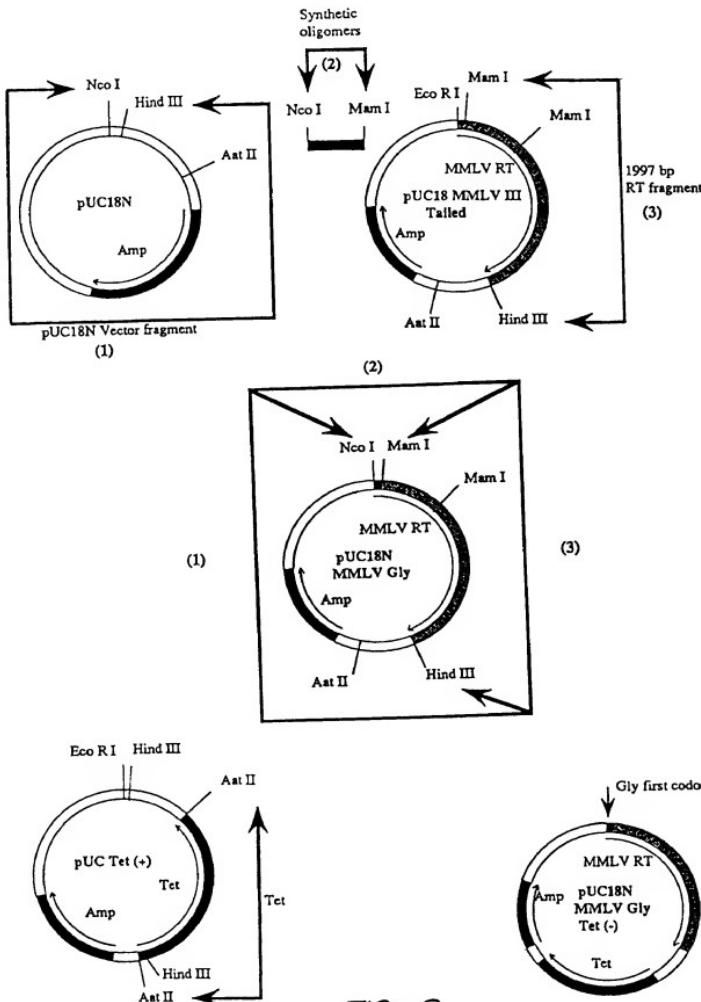


FIG. 8.

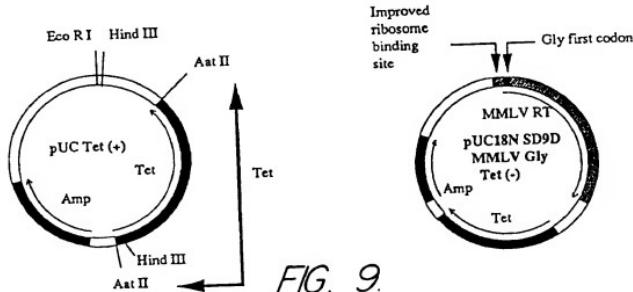
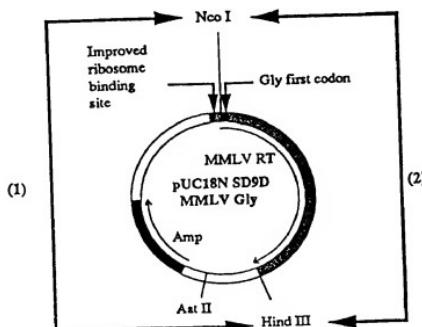
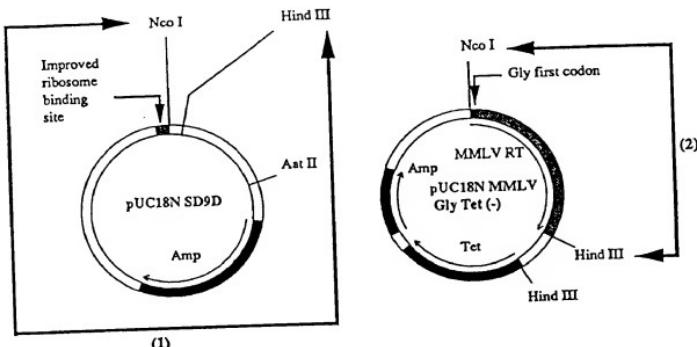
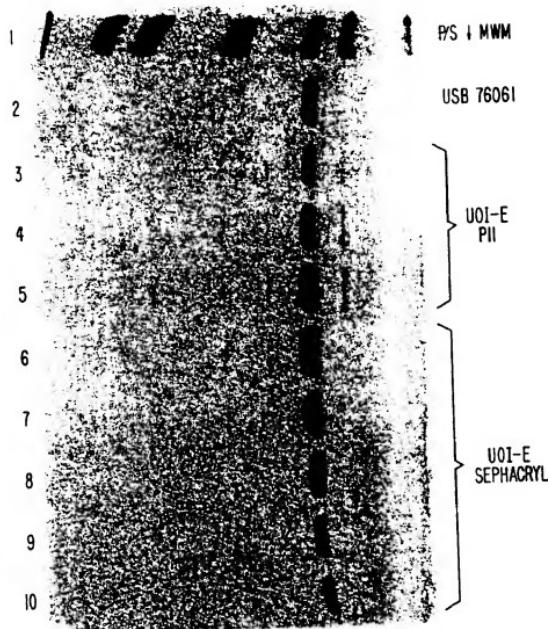


FIG. 9.

FIG. 10





European Patent  
Office

EUROPEAN SEARCH REPORT

Application Number  
EP 95 10 4959

DOCUMENTS CONSIDERED TO BE RELEVANT								
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim						
X,D	US,A,5 017 492 (KOTEWICZ MICHAEL L ET AL) * the whole document ---	1-52						
X,D	US,A,4 943 531 (GOFF STEPHEN P ET AL) * the whole document ---	1-52						
Y	PROC. NATL. ACAD. SCI. USA, vol. 76, no. 11, November 1979 pages 5774-5778, HÄUTALA J. A. ET AL. 'Increased expression of a eukaryotic gene in Escherichia coli through stabilization of its messenger RNA' * abstract ---	1-52						
Y	J. BIOL. CHEM., vol. 267, no. 23, 15 August 1992 pages 16015-16018, KELLY K. O. ET AL. 'RNase PH is essential for the tRNA processing and viability in RNase-deficient Escherichia coli cells' * page 16015, column 1 - column 2 * -----	1-52						
		TECHNICAL FIELDS SEARCHED (INCLUDE)						
		C12N C07K						
<p>The present search report has been drawn up for all claims</p> <hr/> <table border="1"> <tr> <td>Place of search</td> <td>Date of completion of the search</td> <td>Examiner</td> </tr> <tr> <td>MUNICH</td> <td>24 August 1995</td> <td>Chakravarty, A</td> </tr> </table>			Place of search	Date of completion of the search	Examiner	MUNICH	24 August 1995	Chakravarty, A
Place of search	Date of completion of the search	Examiner						
MUNICH	24 August 1995	Chakravarty, A						
<b>CATEGORY OF CITED DOCUMENTS</b>								
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background D : non-critical disclosure P : intermediate document								
T : theory or principle underlying the invention E : earlier patent document, but published on, or prior to, the filing date of the application D : document cited in the application L : document cited for other reasons A : member of the same patent family, corresponding document								